

Alogaritma paralel regularized markov clustering pada jaringan interaksi protein-protein menggunakan format data sparse ELLPACK-R = Balanced and scalable markov clustering of protein interaction networks using ELLPACK-R sparse data format on GPU computing parallel regularized markov clustering algorithm on protein-protein interaction networks using ELLPACK-R sparse data format

Umbu Maramba Mesa, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20308103&lokasi=lokal>

Abstrak

Pada jaringan interaksi protein-protein terdapat beberapa protein yang menjadi pusat cluster, dimana protein-protein tersebut merupakan protein yang memegang peranan penting dalam sebuah fungsi seluler. Salah satu algoritma yang dewasa ini sering digunakan untuk melakukan pencarian pusat kluster adalah algoritma Markov Clustering (MCL). Algoritma Regularized Markov Clustering (R-MCL) merupakan algoritma modifikasi MCL yang bertujuan untuk mencari pusat kluster dengan mensimulasikan random walk dalam graf interaksi protein-protein dengan menggunakan operasi ekspansi namun tetap mempertahankan topologi awal dari graf. Komputasi paralel diperlukan dalam menyelesaikan proses klusterisasi ini sebab R-MCL melibatkan data yang berukuran besar dan mengandung proses yang memiliki kompleksitas waktu yang besar. Dalam skripsi ini akan dibahas mengenai konstruksi algoritma paralel R-MCL menggunakan bahasa pemrograman CUDA C pada GPU. Data disimpan dalam format yang lebih hemat memori yaitu format data sparse ELLPACK-R yang sesuai untuk komputasi pada GPU. Algoritma paralel ini akan diimplementasikan pada mesin manycore dengan menggunakan NVCC compiler.

<hr>

There are some proteins in protein-protein interaction network that act as the cluster centers because of the important roles they have related to cellular functions. One of the clustering algorithms that are often used in clustering is Markov Clustering Algorithm (MCL). Regularized Markov Clustering (R-MCL) algorithm is a modification of MCL in order to get better results by simulating random walk in the graph using expansion operation while maintaining the original topology of the graph. Parallel computation is needed to solve this clustering problem because R-MCL algorithm uses a big number of data and contains some operations with very big time complexities. The problem that will be discussed in this minithesis is the construction of parallel R-MCL algorithm using CUDA C on GPU. The PPI data will be converted into a more memory-friendly format, in this case in ELLPACK-R sparse data format that is suitable for GPU computation. This parallel algorithm will be implemented using a manycore machine with NVCC compiler installed on it.